

1/3/2 (Item 2 from file: 351) [Links](#)  
Fulltext available through: [Order File History](#)  
Derwent WPI  
(c) 2010 Thomson Reuters. All rights reserved.

0012939627 *Drawing available*  
WPI Acc no: 2003-016275/200301

XRPX Acc No: N2003-012195

**Information processing method involves converting speech information to character data which is subjected to prescribed processing based on information expressing characteristic of speech information**

Patent Assignee: OHDAIRA T (OHDA-I); SONY COMPUTER ENTERTAINMENT KK (SONY); YAMAMOTO H (YAMA-I)

Inventor: OHDAIRA T; OHIRA T; YAMAMOTO H

Patent Family ( 2 patents, 2 countries )

Patent Number	Kind	Date	Application Number	Kind	Date	Update	Type
US 20020111794	A1	20020815	US 200275000	A	20020213	200301	B
JP 2002244688	A	20020830	JP 200138224	A	20010215	200301	E

Priority Applications (no., kind, date): JP 200138224 A 20010215

Patent Details

Patent Number	Kind	Lan	Pgs	Draw	Filing Notes
US 20020111794	A1	EN	23	11	
JP 2002244688	A	JA	18		

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-244688  
(43)Date of publication of application : 30.08.2002

(51)Int.Cl. G10L 13/00  
G06F 17/21  
G06F 17/22  
G10L 13/08  
G10L 15/00

(21) Application number : 2001-038224

(71)Applicant : SONY COMPUTER  
ENTERTAINMENT INC

(22) Date of filing : 15.02.2001

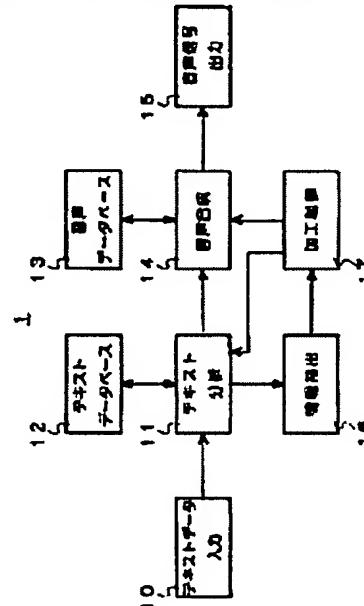
(72)Inventor : YAMAMOTO HIROSHI  
OHIRA TOSHIMITSU

(54) INFORMATION PROCESSOR, INFORMATION PROCESSING METHOD, INFORMATION TRANSMISSION SYSTEM, MEDIUM FOR MAKING INFORMATION PROCESSOR RUN INFORMATION PROCESSING PROGRAM, AND INFORMATION PROCESSING PROGRAM

**(57)Abstract:**

**PROBLEM TO BE SOLVED:** To actualize information processing for rich and joyful expression with, for example, an added feeling and to actualize smooth communication without increasing the amount of transmitted information.

**SOLUTION:** A text analysis part 11 and a voice synthesis part 14 synthesize a voice signal from input text data. An information extraction part 16 extracts, for example, a stop mark and punctuation marks from the input text data. A processing control part 17 controls the text analysis part 11 and voice synthesis part 14 according to the extraction of the stop mark and punctuation marks to add a word or phrase expressing complaining anger or a word or phrase showing a light laugh to the ending of words.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2002-244688  
(P2002-244688A)

(43)公開日 平成14年8月30日(2002.8.30)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード*(参考)
G 1 0 L 13/00		G 0 6 F 17/21	5 6 8 Z 5 B 0 0 9
G 0 6 F 17/21	5 6 8	17/22	5 0 3 5 D 0 1 5
17/22	5 0 3	G 1 0 L 3/00	E 5 D 0 4 5
G 1 0 L 13/08			H
15/00			5 5 1 A

審査請求 未請求 請求項の数34 O.L (全 18 頁)

(21)出願番号 特願2001-38224(P2001-38224)

(71)出願人 395015319

株式会社ソニー・コンピュータエンタテインメント  
東京都港区赤坂7-1-1

(22)出願日 平成13年2月15日(2001.2.15)

(72)発明者 山本 浩  
東京都港区赤坂7丁目1番1号 株式会社  
ソニー・コンピュータエンタテインメント  
内

(74)代理人 100107238

弁理士 米山 尚志

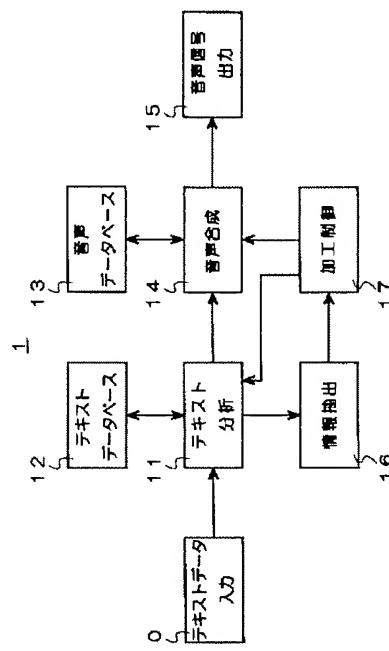
最終頁に続く

(54)【発明の名称】 情報処理方法及び装置、情報伝送システム、情報処理プログラムを情報処理装置に実行させる媒体、情報処理プログラム

(57)【要約】

【課題】 例えば感情を加味した豊かで楽しい表現を可能とする情報処理を実現し、また、情報伝達を行う場合に、伝送情報量を増やすことなく、円滑なコミュニケーションを実現する。

【解決手段】 テキスト分析部11と音声合成部14では、入力テキストデータから音声信号を合成する。また、情報抽出部16では、入力テキストデータから例えば句点や読点等を抽出する。加工制御部17では、その句点や読点の抽出に応じて、テキスト分析部11や音声合成部14を制御することにより、例えば語尾に文句的な怒りを表現する語句や、軽い笑いの語句を付加することなどを行う。



1

## 【特許請求の範囲】

【請求項1】 ネットワークを介して文字データを受信し、

上記受信した文字データから所定の情報を抽出し、

上記文字データを他の情報へ変換し、

上記抽出した所定の情報に基づいて、上記文字データ又は上記他の情報に所定の加工処理を施すことを特徴とする情報処理方法。

【請求項2】 上記所定の情報として、上記文字データから所定の文字コードを抽出することを特徴とする請求項1記載の情報処理方法。

【請求項3】 上記所定の情報として、上記文字データから所定の単語又は語句を抽出することを特徴とする請求項1記載の情報処理方法。

【請求項4】 上記他の情報として、上記文字データを音声情報へ変換することを特徴とする請求項1から請求項3のうち、いずれか一項記載の情報処理方法。

【請求項5】 上記所定の加工処理として、上記文字データ若しくは上記他の情報に対して、感情を表現する語を付加、若しくは感情を表現する語への変換を行うことを特徴とする請求項1から請求項4のうち、いずれか一項記載の情報処理方法。

【請求項6】 入力情報から当該入力情報の特徴を現す情報を抽出し、

上記入力情報を文字データへ変換し、

上記抽出した特徴を現す情報に基づいて、上記文字データに所定の変換処理を施し、

上記所定の変換処理が施された文字データをネットワークを介して送信することを特徴とする情報処理方法。

【請求項7】 上記入力情報としての音声情報を上記文字データへ変換することを特徴とする請求項6記載の情報処理方法。

【請求項8】 上記所定の変換処理として、上記文字データの文字態様を変更することを特徴とする請求項6又は請求項7記載の情報処理方法。

【請求項9】 上記所定の変換処理として、上記文字データの制御コードを変更することを特徴とする請求項6又は請求項7記載の情報処理方法。

【請求項10】 上記入力情報の特徴を現す情報として、感情を現す情報を抽出することを特徴とする請求項6から請求項9のうち、いずれか一項記載の情報処理方法。

【請求項11】 ネットワークを介して送信されてきた文字データを受信する受信手段と、

上記受信した文字データから所定の情報を抽出する情報抽出手段と、

上記文字データを他の情報へ変換する変換手段と、

上記抽出した所定の情報に基づいて、上記文字データ又は上記他の情報に所定の加工処理を施す加工処理手段とを有することを特徴とする情報処理装置。

10

【請求項12】 上記情報抽出手段は、上記所定の情報として、上記文字データから所定の文字コードを抽出することを特徴とする請求項1記載の情報処理装置。

【請求項13】 上記情報抽出手段は、上記所定の情報として、上記文字データから所定の単語又は語句を抽出することを特徴とする請求項1記載の情報処理装置。

【請求項14】 上記変換手段は、上記他の情報として、上記文字データを音声情報へ変換することを特徴とする請求項1から請求項3のうち、いずれか一項記載の情報処理装置。

【請求項15】 上記加工処理手段は、上記所定の加工処理として、上記文字データ若しくは上記他の情報に対して、感情を表現する語を付加、若しくは感情を表現する語への変換を行うことを特徴とする請求項1から請求項4のうち、いずれか一項記載の情報処理装置。

【請求項16】 入力情報から当該入力情報の特徴を現す情報を抽出する情報抽出手段と、

上記入力情報を文字データへ変換する入力情報変換手段と、

20

上記抽出した特徴を現す情報に基づいて上記文字データに所定の変換処理を施す文字データ変換手段と、

上記所定の変換処理が施された文字データをネットワークを介して送信する送信手段とを有することを特徴とする情報処理装置。

【請求項17】 上記入力情報変換手段は、上記入力情報としての音声情報を上記文字データへ変換することを特徴とする請求項16記載の情報処理装置。

【請求項18】 上記文字データ変換手段は、上記所定の変換処理として、上記文字データの文字態様を変更することを特徴とする請求項16又は請求項17記載の情報処理装置。

30

【請求項19】 上記文字データ変換手段は、上記所定の変換処理として、上記文字データの制御コードを変更することを特徴とする請求項16又は請求項17記載の情報処理装置法。

【請求項20】 上記情報抽出手段は、上記入力情報の特徴を現す情報として、感情を現す情報を抽出することを特徴とする請求項16から請求項19のうち、いずれか一項記載の情報処理装置。

40

【請求項21】 入力情報から当該入力情報の特徴を現す情報を抽出する特徴情報抽出手段と、

上記入力情報を文字データへ変換する入力情報変換手段と、

上記抽出した特徴を現す情報に基づいて上記文字データに所定の変換処理を施す文字データ変換手段と、

上記所定の変換処理が施された文字データをネットワークを介して送信する送信手段とを有する第1の情報処理装置と、

50

ネットワークを介して送信されてきた文字データを受信する受信手段と、

上記受信した文字データから所定の情報を抽出する所定情報抽出手段と、

上記文字データを他の情報へ変換する変換手段と、

上記抽出した所定の情報に基づいて、上記文字データを他の情報へ変換する変換手段と、

上記抽出した所定の情報を基づいて、

て上記文字データ又は上記他の情報に所定の加工処理を施す加工処理手段とを有する第2の情報処理装置とを備えることを特徴とする情報伝送システム。

【請求項22】 ネットワークを介して送信されてきた文字データを受信するステップと、  
上記受信した文字データから所定の情報を抽出するステップと、  
上記文字データを他の情報へ変換するステップと、  
上記抽出した所定の情報に基づいて、上記文字データ又は上記他の情報に所定の加工処理を施すステップとを含むことを特徴とする情報処理プログラムを情報処理装置に実行させる媒体。

【請求項23】 上記所定の情報として、上記文字データから所定の文字コードを抽出するステップを含むことを特徴とする請求項22記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項24】 上記所定の情報として、上記文字データから所定の単語又は語句を抽出するステップを含むことを特徴とする請求項22記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項25】 上記他の情報として、上記文字データを音声情報へ変換するステップを含むことを特徴とする請求項22から請求項24のうち、いずれか一項記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項26】 上記所定の加工処理として、上記文字データ若しくは上記他の情報に対して、感情を表現する語を付加、若しくは感情を表現する語への変換を行うステップを含むことを特徴とする請求項22から請求項23のうち、いずれか一項記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項27】 入力情報から当該入力情報の特徴を現す情報を抽出するステップと、  
上記入力情報を文字データへ変換するステップと、  
上記抽出した特徴を現す情報に基づいて、上記文字データに所定の変換処理を施すステップと、  
上記所定の変換処理が施された文字データをネットワークを介して送信するステップとを含むことを特徴とする情報処理プログラムを情報処理装置に実行させる媒体。

【請求項28】 上記入力情報としての音声情報を上記文字データへ変換するステップを含むことを特徴とする請求項27記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項29】 上記所定の変換処理として、上記文字データの文字態様を変更するステップを含むことを特徴とする請求項27又は請求項28記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項30】 上記所定の変換処理として、上記文字データの制御コードを変更するステップを含むことを特徴とする請求項27又は請求項28記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項31】 上記入力情報の特徴を現す情報として、感情を現す情報を抽出するステップを含むことを特徴とする請求項27から請求項30のうち、いずれか一項記載の情報処理プログラムを情報処理装置に実行させる媒体。

【請求項32】 所定の記録媒体又は伝送媒体を含むことを特徴とする請求項27から請求項31のうち、いずれか一項記載の情報処理処理プログラムを情報処理装置に実行させる媒体。

【請求項33】 ネットワークを介して送信されてきた文字データを受信するステップと、  
上記受信した文字データから所定の情報を抽出するステップと、  
上記文字データを他の情報へ変換するステップと、  
上記抽出した所定の情報に基づいて、上記文字データ又は上記他の情報に所定の加工処理を施すステップとを含むことを特徴とする情報処理プログラム。

【請求項34】 入力情報から当該入力情報の特徴を現す情報を抽出するステップと、

上記入力情報を文字データへ変換するステップと、  
上記抽出した特徴を現す情報に基づいて、上記文字データへの変換処理を制御するステップと、  
上記所定の変換処理が施された文字データをネットワークを介して送信するステップとを含むことを特徴とする情報処理プログラム。

【発明の詳細な説明】

#### 【0001】

【発明の属する技術分野】 本発明は、例えば音声情報等から文字データへの情報変換、若しくは文字データから音声情報等への情報変換処理や、例えば文字データに附加された情報基づいて音声情報等を加工するような情報加工処理を行う情報処理方法及び装置、例えばテキスト情報を伝送する情報伝送システム、情報処理プログラムを情報処理装置に実行させる媒体、情報処理プログラムに関する。

#### 【0002】

【従来の技術】 従来より、テキスト情報を音声に変換したり、音声をテキスト情報へ変換するようなテキスト音声変換システムが存在している。

【0003】 すなわち、上記テキスト音声変換システムでは、例えば、以下のテキスト音声合成処理によりテキスト情報から音声への変換を行い、一方、以下の音声認識処理により音声からテキスト情報への変換を行う。

【0004】 上記テキスト音声合成処理は、大別して、言語処理と音響処理に分けられる。

【0005】 上記言語処理では、入力されたテキスト情報（例えば漢字や仮名からなる文の情報）を、単語の読みとアクセント、文のイントネーションに関する情報を保持する表音文字列へ変換する。具体的に説明すると、

当該言語処理では、予め用意されている単語辞書を用い

て、入力テキスト文中の各単語の読みとアクセントを決定し、さらに、各文節の係り受け関係（修飾句・節の中に更に修飾節がかかるような文の関係）から、文章のイントネーションを決めるこことにより、上記テキスト情報を表音文字列へ変換する。

【0006】上記音響処理では、予め用意されている波形辞書を用いて、上記表音文字列を構成する各音素の波形を読み出し、音声波形（音声信号）を組み上げる。

【0007】上述のように、上記テキスト音声変換システムでは、上記テキスト音声合成処理による音声波形（音声信号）が、上記テキスト情報から音声への変換結果として得られることになる。

【0008】一方、上記音声認識処理は、大別して、音声入力処理、周波数分析処理、音素認識処理、単語認識処理、文認識処理に分けられる。

【0009】上記音声入力処理では、例えばマイクロホン等により、音声を電気信号（音声信号）に変換する。

【0010】上記周波数分析処理では、上記音声入力処理により得られた音声信号を、例えば数ミリ秒～十数ミリ秒毎のフレームに分割し、さらに各フレーム毎にスペクトル分析を行う。なお、上記スペクトル分析には、高速フーリエ変換などが用いられる。上記スペクトル分析により得られたフレーム毎のスペクトル成分は、雑音成分が取り除かれ、さらに人間の聴覚尺度に基づく音声パラメータに変換される。

【0011】上記音素認識処理では、上記周波数分析処理により得られた時系列の音声パラメータと、予め用意されている音素モデルとの照合を行い、その照合により得られた音素モデルから音素が求められる。すなわち、音素、特に子音の特徴は音声スペクトルの時間変化のパターンとして現れるため、当該音素認識処理では、上記音声パラメータを時系列で表現した音素モデルと、上記周波数分析処理により得られた時系列の音声パラメータとの照合を行い、その照合により得られた音素モデルから音素を求める。なお、上記音素モデルは、多数の音声パラメータから予め学習により求められている。当該学習には、例えば、時系列パターンをマルコフモデル化した、いわゆる隠れマルコフモデル（HMM：Hidden Markov Model）などが用いられる。

【0012】上記単語認識処理では、上記音素認識処理により得られた音素認識結果と、予め用意されている単語モデルとの照合を行い、両者の一致度を計算し、その一致度が最も高い単語モデルから単語を求める。ここで、単語モデルとしては、単語中の母音の無声化・長音化・鼻音化、子音の口蓋化など、音素の変形を考慮したモデルが用意される。なお、各音素の発声タイミングの変動については、いわゆる動的計画法の原理を用いた照合法（D Pマッチング）で対処するのが一般的となっている。

【0013】上記文認識処理では、上記単語認識処理に

より得られた単語認識結果から、言語モデル（単語と単語のつながりのモデル或いは文法）に合致する単語列を選び出す。

【0014】上記テキスト音声変換システムでは、上記音声認識処理による上記単語列からなるテキスト情報が、音声からテキスト情報への変換結果として得られることになる。

【0015】また、上述のようなテキスト音声変換システムは、ネットワークを介した情報伝達への応用も検討されており、例えば入力音声から変換されたテキスト情報をネットワークを介して伝送するような情報伝達システム、或いは、伝送されてきたテキスト情報（例えば電子メール等）を音声に変換して出力するような情報伝送システムも既に考え出されている。

#### 【0016】

【発明が解決しようとする課題】ところで、上記テキスト音声変換システムにおいて、上記テキスト音声合成処理によるテキスト情報から音声への変換や、上記音声認識処理による音声からテキスト情報への変換の際には、間違いない正確な変換を行うことが要求されている。

【0017】このため、上記テキスト音声合成処理により得られる音声は、正確ではあるが機械的であり、例えば人間が話すときの音声のように感情が伴った音声ではなく、非人間的な音声になることが多い。同様に、上記音声認識処理により得られるテキスト情報は、正確ではあるが、話者の感情が伴った音声を表現することまではできない。

【0018】また、例えば上記テキスト音声変換システムとネットワークを介した情報伝達とを組み合わせて情報伝送システムを構築することを考えた場合、従来のテキスト音声変換システムでは、上述のように変換後のテキスト情報や音声が人間の感情を表現できないため、送受信側双方で感情を含めた意思の疎通を図ることが困難であり、不要な誤解を生む虞もある。

【0019】なお、上記音声を変換したテキスト情報と共に、当該音声を同時に伝達する（例えば音声を圧縮した圧縮データファイルをテキストデータの添付ファイルとして送信する）ことで、受信側へ送信側の感情をそのまま伝えるようなことも考えられるが、この場合は伝送情報量が多くなり好ましくない。

【0020】また、上述のように、受信側に対してテキスト情報と共に音声データ（音声圧縮データ）を送信するようにした場合、当該受信側へ送られる音声圧縮データは、送信側の音声そのものであり、送信側の感情がダイレクト且つリアルに受信側へ伝えられることになるため、逆に好ましくない場合もあり得る。すなわち、送受信側双方のコミュニケーションを円滑に進めるために、送信側から受信側へ伝えられる感情表現はリアル過ぎるよりもある程度和らげられた表現である方が好ましい。さらに一步進んで、楽しい感情表現や、誇張した感

情表現などができるば、コミュニケーションを円滑にする上では、より望ましいと考えられる。

【0021】そこで、本発明は、このような課題に鑑みてなされたものであり、例えば感情を伴った豊かで楽しい表現を可能とする情報交換を実現し、また、例えば情報伝達を行う場合において、伝送情報量を増やすことなく、円滑なコミュニケーションを実現可能とする、情報処理方法及び装置、情報伝送システム、情報処理プログラムを情報処理装置に実行させる媒体、情報処理プログラムを提供することを目的とする。

#### 【0022】

【課題を解決するための手段】本発明は、ネットワークを介して受信した文字データから所定の情報を抽出すると共に当該文字データを他の情報へ変換し、抽出した所定の情報に基づいて文字データ又は他の情報に所定の加工処理を施すようにしている。

【0023】すなわち、本発明によれば、文字データに元々含まれている所定の情報に基づいて所定の加工処理を行うよう正在しているため、所定の加工処理の際に、特別な情報を用意する必要がない。

【0024】また、本発明は、入力情報の特徴を現す情報を抽出すると共にその入力情報を文字データへ変換し、当該抽出した特徴を現す情報を基づいて文字データに所定の変換処理を施してネットワークへ送信するよう正在している。

【0025】すなわち、本発明によれば、入力情報から抽出された当該入力情報の特徴を現す情報に基づいて文字データに変換処理を施すよう正在しているため、当該変換処理後の文字データは、上記特徴を現す情報が明示的に付加された状態の情報となり、また上記特徴を現す情報が付加されたとしても情報量の増加は殆ど無い。

#### 【0026】

【発明の実施の形態】【第1の実施の形態の情報処理装置】図1には、本発明の第1の実施の形態として、入力文字データ（以下、テキストデータとする）を音声信号へ変換する情報処理装置1の概略構成を示す。なお、この図1の構成は、ハードウェア構成、或いはソフトウェアの何れでも実現可能である。

【0027】この図1において、テキストデータ入力部10には、例えばインターネットやイーサネット（登録商標）等のネットワークの通信回線を介して伝送されたテキストデータ（電子メール等を含む）や、例えばキーボード等により入力されたテキストデータ、記録媒体から再生されたテキストデータなどが供給され、テキスト分析部11へ送られる。

【0028】テキスト分析部11は、テキストデータベース部12に予め用意されている単語辞書を用いて、入力テキストデータ中の各単語の読みとアクセントを決定し、さらに、各文節の係り受け関係から、文章のインтоネーションを決めるこにより、上記テキストデータを

表音文字列へ変換する。また、当該テキスト分析部11では、必要に応じて上記入力テキストデータを所望の言語に変換（翻訳）し、当該変換された言語のテキストデータから上記表音文字列への変換を行ふことも可能となつてゐる。当該テキスト分析部11により得られた表音文字列のデータは、音声合成部14へ送られる。

【0029】音声合成部14は、音声データベース部13に予め用意されている波形辞書を用いて、上記表音文字列を構成する各音素の波形を読み出し、音声波形（音声信号）を組み上げる。

【0030】以上説明した上記テキストデータ入力部10、テキスト分析部11、音声合成部14における各処理は、前述した従来のテキスト音声変換システムのテキスト音声合成処理と略々同様の処理である。もちろん、本発明におけるテキストデータから音声信号への変換処理は上述の例に限定されるものではなく、他のテキスト音声変換処理方法を用いても良い。

【0031】一方で、本発明実施の形態の情報処理装置1は、上記入力テキストデータに含まれる所定の情報に基づいて、上記音声合成時若しくはテキスト分析の際の表音文字列の生成時に、例えば感情、思考、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを加味した合成音声若しくは表音文字列を生成するような情報加工を行ふようになされている。

【0032】すなわち、本実施の形態の情報処理装置1は、上記入力テキストデータに含まれる所定の情報に基づいて、上記音声合成時若しくはテキスト分析時に情報加工処理を行ふための構成として、情報抽出部16と加工制御部17とを備えている。

【0033】上記情報抽出部16は、上記テキスト分析部11が入力テキストデータを分析して得られた文字コード（制御コードやASCⅡ文字、片仮名、漢字、補助漢字等を含む文字コード）のうちの所定文字コードや、テキスト文に付随するヘッダやフッダ、テキスト文内の所定の語句や単語を、上記所定の情報として抽出し、その抽出した所定の情報を加工制御部17へ送る。

【0034】上記加工制御部17は、上記所定の情報に基づいて、上記音声合成部14での音声合成処理や上記テキスト分析部11でのテキスト分析処理の際に、例えば感情、思考、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを加味した合成音声若しくは表音文字列を生成させるような加工制御を行う。

【0035】ここで、本実施の形態の情報処理装置1において、上記情報抽出部16が上記入力テキストデータから抽出する所定の情報としては、例えば、テキストの文字の太さ、文字の大きさ、文字の色、文字種類、文字の配置位置、テキスト文のスタイル、体裁、注釈、句点、句読点、読点などを現す各種のコードや、テキストデータに付加されているヘッダ、フッダ、テキスト文中に含まれる語句や単語などを挙げることができる。

【0036】また、本実施の形態において、上記情報抽出部16が上記テキストデータから抽出する所定の情報と、その所定の情報に応じて上記加工制御部17が行う加工制御との対応としては、例えば、テキストデータから上記所定の情報として抽出された太い文字に応じて話者の感情の高まりや怒りを表現する合成音声等が得られるようにする加工制御、逆に、細い文字に応じて話者の感情の落ち込みや悲しみを表現する合成音声等が得られるようにする加工制御、例えば大きいサイズの文字で大人を表現する合成音声等が得られるようにする加工制御、小さいサイズの文字で子供を表現する合成音声等が得られるようにする加工制御、青色の文字により性別の男性を表現する合成音声等が得られるようにする加工制御、ピンク色の文字により性別の女性を表現する合成音声等が得られるようにする加工制御などが考えられる。

【0037】また、本実施の形態において、テキストデータから抽出される語句や単語と、上記加工制御との対応としては、例えば「ボリューム大」や「感情高い」、「テンポ速い」のような語句により感情の高まり等を表現する合成音声等を得るような加工制御や、「ボリューム小」や「感情低い」、「テンポ遅い」のような語句により感情の低下を等を表現する合成音声等を得るような加工制御などが考えられる。

【0038】さらに、本実施の形態では、テキストデータから抽出した句点や句読点、読点に応じて、任意の言葉や単語を追加、変更、付加などしたり、語尾を加工するなどの加工制御なども考えられる。なお、本実施の形態において、上記語尾を加工する制御としては、例えば「そうだ(ニヤー)」、「そうだ(ワン)」などのように語句の最後に「ニヤー」や「ワン」など猫や犬に対する嗜好を表現する語句を自動的に挿入したり、また、例えば「明日の会議ですが、諸々の事情により、延期したいと思います」のような会議調の言葉に対して「明日の会議ですが(え~)、諸々の事情により(え~)、延期したいと思います(え~)」など通常の会話時に思考中であることを現す「え~」のような語を、読点の後などに自動的に挿入するような加工制御や、例えば「まったくさあ」、「どういうつもりだよ」、「ホントにもう」などの文句的な怒りを表現する単語を句点の後などに付加したり、例えば「あはは」、「くすぐす」、「うふふ」など楽しさを表現する軽い笑いの語句を句読点の後に付加するような加工制御が考えられる。

【0039】その他、本実施の形態では、例えば単語を変更するような加工制御として、例えば「そうですね」を「そうでんねん」に変換したり、「こんにちは」を「Hello」や「Guten Tag」、「ニーハオ」、「Bonjour」に変換するように、方言変換や言語変換(翻訳)などの加工の制御が考えられる。

【0040】なお、上述した各所定の情報と加工制御との対応付けは一例であり、上記の例に限定されるもので

はなく、それらの対応付けの組み合わせはシステムで任意に設定可能である。

【0041】このように、本実施の形態の情報処理装置1によれば、テキストデータの文字コードやヘッダ、フッダ、語句や単語などの所定の情報に応じて、音声合成処理やテキスト分析処理の際に、感情、思考、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを加味した情報加工処理を行うことが可能となっている。また、本実施の形態によれば、上記所定の情報は、文字コードや単語、語句等のように、テキストデータを構成する情報そのものが用いられているため、情報加工制御のために特別な情報を扱う必要がなく、専用のソフトウェアなども不要である。

【0042】本実施の形態の情報処理装置1では、上記音声合成部14により合成された合成音声信号が、音声信号出力部15から図示しない後段の構成へ出力される。なお、実際に放音を行う場合、上記音声信号出力部15は、スピーカ等の電気音響変換手段を含むことになる。

【0043】本実施の形態では、上記テキストデータを合成音声に変換して出力する例を挙げたが、上述のように加工処理を施した後のテキスト文を例えばモニタ装置等の画面上にそのまま表示することも可能である。このように加工処理後のテキスト文を画面表示する場合、例えば聴覚障害を持った人でもその内容を認識することが可能となる。

【0044】[第2の実施の形態の情報処理装置]次に、図2には、本発明の第2の実施の形態として、入力音声信号をテキストデータへ変換する情報処理装置2の概略構成を示す。なお、この図2の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。

【0045】この図2において、音声信号入力部21には、例えばマイクロホン等の音響電気変換素子により得られた音声信号や、通信回線を介して伝送されてきた音声信号、記録媒体から再生された音声信号などが供給され、当該入力音声信号は音声分析部22へ送られる。

【0046】音声分析部22は、上記音声信号入力部21から送られてきた音声信号のレベル分析を行うと共に、当該音声信号を例えば数ミリ秒～十数ミリ秒毎にフレーム化し、さらに各フレーム毎に例えば高速フーリエ変換等によるスペクトル分析を行う。当該スペクトル分析により得られたフレーム毎のスペクトル成分は、雑音成分が取り除かれ、さらに人間の聴覚尺度に基づく音声パラメータに変換されて音声認識部23へ送られる。

【0047】上記音声認識部23は、上記音声分析部22から供給された時系列の音声パラメータと、音声データベース部24に予め用意されている音素モデルとの照合を行い、その照合により得られた音素モデルから音素を求める、得られた音素認識結果をテキスト変換部26へ

送る。なお、上記音声データベース部24に用意されている音素モデルは、例えば、前記隠れマルコフモデル(HMM)などを用いた学習により予め求められたものである。

【0048】上記音声認識部23から音素認識結果を受け取ったテキスト変換部26は、テキストデータベース部25に予め用意されている単語モデルと、上記音素認識結果との照合を行って両者の一致度を計算し、その一致度が最も高い単語モデルから単語を求め、さらに、同じくテキストデータベース部25に予め用意されている言語モデルと上記単語の認識結果との紹介を行って合致する単語列を選び出してテキストデータを生成する。なお、上記テキストデータベース部25に用意されている単語モデルは、単語中の母音の無声化、長音化、鼻音化、子音の口蓋化など、音素の変形を考慮したモデルとして予め求められているものであり、また、上記言語モデルは、単語と単語のつながりのモデル或いは文法として予め求められているものである。

【0049】以上説明した上記音声信号入力部21、音声分析部22、音声認識部23、テキスト変換部26における各処理は、前述した従来のテキスト音声変換システムの音声認識処理と略々同様の処理である。もちろん、本発明における音声信号からテキストデータへの変換処理は上述の例に限定されるものではなく、他の音声テキスト変換処理方法を用いても良い。

【0050】一方で、本発明の第2の実施の形態の情報処理装置2では、上記入力音声信号から話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定し、それら特定結果に応じて上記テキスト変換部26でのテキスト変換処理を制御することにより、例えば後に、前記第1の実施の形態において当該テキストデータを分析して音声合成を行われることになった場合に、前記テキスト分析(翻訳等の言語変換を含む)や音声合成時に、上記感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを加味した加工処理を行えるようにしている。なお、言うまでもないが、前記第1の実施の形態の情報処理装置1は、当該第2の実施の形態の情報処理装置2によるテキスト変換処理が施されていない一般的なテキストデータであっても、前述したような加工処理を行えるものである。

【0051】本実施の形態の情報処理装置2は、上記入力音声信号から上記感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定し、その特定結果に基づいて上記テキスト変換処理を制御するための構成として、テキスト変換制御部29と声紋・特徴データベース部30とを備えている。

【0052】上記テキスト変換制御部29は、上記音声分析部22での分析処理による上記入力音声信号のスペクトル成分やレベル成分、上記テキスト変換部26にお

いて上記音素認識結果から変換されたテキストデータなどに基づいて、上記入力音声信号に含まれる感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定し、その特定結果に応じた制御コマンドを上記テキスト変換部26へ送る。

【0053】すなわち、当該テキスト変換制御部29では、例えば、いわゆる声紋分析理論に基づき、上記入力音声信号のスペクトル成分やレベル成分と、声紋・特徴データベース30に予め用意されている声紋を現す特徴データとを照合することにより、話者の感情や思考、声帯、口腔、鼻腔の形状や顔の骨格(つまり顔形)、身体全体の骨格や身長、体重、性別、年齢、職業、出身地、例えば風邪を引いているときのくしゃみや咳による体調等を特定する。また、当該テキスト変換制御部29では、例えば上記音声分析部22による分析結果や上記テキスト変換部26で上記音素認識結果から変換されたテキストデータと、上記声紋・特徴データベース30に予め用意されている職業、出身地、趣味、嗜好などを特徴的に現す特徴データとを照合することにより、話者の職業、出身地、趣味、嗜好などを特定する。そして、テキスト変換制御部29は、上記特定された話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などをに基づいて、上記テキスト変換部26において上記音素認識結果から変換されたテキストデータに付加する文字コードや変更する文字コード、或いは、ヘッダやフッダ、単語、語句などを決定し、その決定に応じた制御コマンドをテキスト変換部26へ送る。

【0054】ここで、本実施の形態において、上記テキスト変換制御部29が上記テキスト変換部26へ送る制御コマンドは、例えば、上記テキスト変換部26が上記音素認識結果から変換したテキストデータに対して、例えば文字の大きさ、サイズ(フォントサイズ)、文字の色、文字種類(字体、仮名漢字、漢数字、アルファベット、アラビア文字、各種記号など)、文字の配置位置(行、列等)、テキスト文のスタイル(字数、行数、行間、字間、マージンなど)、体裁、注釈、句点、句読点、読点などを付加或いは変更させるためのコマンドや、ヘッダ、フッダ、単語、語句などの情報を付加或いは変更させるためのコマンドである。

【0055】すなわち、本実施の形態の情報処理装置2では、例えば前記第1の実施の形態の情報処理装置1にて前記感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを加味した加工処理を行うような場合に、前記テキストデータから抽出した上記文字コードやヘッダ、フッダ、単語、語句等の所定の情報に応じて加工処理ができるようにするために、上記制御コマンドによるテキスト変換を行っている。

【0056】前記第1の実施の形態での加工処理例と対応付けて具体的に説明すると、本実施の形態の情報処理装置2では、例えば上記入力音声信号から話者の感情の

高まりや怒りがあることを特定した場合はそれに対応したテキストの文字を例えば太い文字に変換し、逆に、話者の感情の落ち込みや悲しみを特定した場合はテキストの文字を例えば細い文字に変換したり、話者が大人であることを特定した場合はフォントサイズを大きいサイズに変換したり、子供であることを特定した場合は小さいフォントサイズに変換したり、性別が男性であることを特定した場合は青色の文字に変換したり、性別が女性であることを特定した場合はピンク色の文字に変換したりする。

【0057】また、本実施の形態の情報処理装置2では、上記入力音声信号から話者の感情の高まりや怒りがあることを特定した場合は一例として括弧記号で囲んだ「ボリューム大」や「感情高い」、「テンポ速い」のような語句をテキスト文中に挿入したり、感情の低下を特定した場合は括弧記号で囲んだ例えは「ボリューム小」や「感情低い」、「テンポ遅い」のような語句をテキスト文中に挿入したりする。

【0058】その他、本実施の形態の情報処理装置2では、ヘッダやフッダなどに、例えば語尾の加工や、単語の付加、単語の変更などを要求する情報を挿入することも可能である。

【0059】なお、本実施の形態の情報処理装置2がテキストデータに施す変換処理（つまり前記所定の情報の付加等）と、前記第1の実施の形態の情報処理装置1で行われる加工制御との対応付けは一例であり、上記の例に限定されるものではなく、それらの対応付けの組み合わせはシステムで任意に設定可能である。

【0060】このように、本実施の形態では、前記感情、思考、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを、一般的な文字コードやヘッダ、フッダ等を利用してテキストデータ内に入れ込むようしているため、それら感情等を表現するための情報として新たな情報を用意する必要がない。このため、上記入力音声信号から変換したテキストデータを例えばネットワーク等を介して送信するようなことを考えた場合、前述の従来例で説明した圧縮音声データを送信する場合のように伝送情報量が増えてしまう虞はなく、また、上記感情等を表現するための新規な情報を扱うための専用のソフトウェアなども不要である。

【0061】本実施の形態の情報処理装置2では、上述したようにして入力音声信号から変換されたテキストデータが、テキストデータ出力部27から図示しない後段の構成へ出力される。なお、当該テキストデータをネットワーク上に伝送する場合、上記テキストデータ出力部27は、上記ネットワークへの接続手段を含むことになる。また、上記テキストデータを記録媒体に記録する場合、上記テキストデータ出力部27は、記録媒体への記録手段を含むことになる。

【0062】[第3の実施の形態の情報処理装置] 図3

には、本発明の第3の実施の形態として、入力音声信号と共に例えば話者の画像等も用いてテキスト変換制御を行なうようにした情報処理装置3の概略構成を示す。なお、この図3の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図3の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0063】この第3の実施の形態の情報処理装置3の場合、画像信号入力部31には、例えば音声入力を行っている話者を撮像した画像信号が入力され、その画像信号は画像解析部32へ送られる。

【0064】画像解析部32では、例えば、画像の特徴抽出の一手法である固有空間解析手法等を用いて、例えば話者の顔画像に対するアフィン変換などにより、顔の表情空間を構築して顔の表情を分類し、当該顔表情空間で分類された顔の表情パラメータを取り出し、その表情パラメータをテキスト変換制御部29へ送る。

【0065】当該テキスト変換制御部29では、前記音声分析部22での分析処理による入力音声信号のスペクトル成分やレベル成分、上記テキスト変換部26において前記音素認識結果から変換したテキストデータなどに基づいて、前述の第2の実施の形態の場合と同様に、話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定すると共に、上記表情パラメータを用いて話者の感情や思考、性別、体調、顔形などをさらに特定し、それらの特定結果に応じた前記制御コマンドを生成する。すなわちこの第3の実施の形態の情報処理装置3の場合のテキスト変換制御部29は、前記第2の実施の形態の場合の処理に加えて、画像データベース部33に予め用意されている様々な顔の表情の特徴を現す表情パラメータのパターンと、上記画像解析部32により得られた表情パラメータとを照合し、その照合により話者の感情や思考、性別、体調、顔形などをさらに特定するようしている。具体的には、「楽しい」、「悲しい」、「驚き」、「嫌悪」、「怒り」、「恐れ」などの表情から感情を特定したり、顔の特徴から性別や体調などを特定して制御コマンドを生成する。なお、本実施の形態においても第2の実施の形態の場合と同様に、前記制御コマンドに応じたテキスト変換処理と前記感情等との対応付けは一例であり、当該システムで任意に設定可能であり、上記の例に限定されるものではない。

【0066】このように、本発明の第3の実施の形態によれば、入力音声信号だけでなく、例えば話者の顔画像をも用いているため、話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などをより正確に特定可能となっている。

【0067】[第4の実施の形態の情報処理装置] 図4には、本発明の第4の実施の形態として、入力音声信号

と共に例えば話者の血圧・心拍数等も用いてテキスト変換制御を行なうようにした情報処理装置4の概略構成を示す。なお、この図4の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図4の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0068】この第4の実施の形態の情報処理装置4の場合、血圧・心拍数信号入力部34には、例えば音声入力を行なっている話者に取り付けられら血圧・心拍数測定装置からの測定信号が入力され、その測定信号は血圧・心拍数解析部35へ送られる。

【0069】血圧・心拍数解析部34では、例えば、上記測定信号を解析し、話者の血圧と心拍数を現す血圧・心拍数パラメータを取り出し、そのパラメータをテキスト変換制御部29へ送る。

【0070】当該テキスト変換制御部29では、前記音声分析部22での分析処理による入力音声信号のスペクトル成分やレベル成分、上記テキスト変換部26にて前記音素認識結果から変換されたテキストデータなどに基づいて、前述同様に、話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定すると共に、上記血圧・心拍数パラメータを用いてさらに詳細な特定を行い、それらの特定結果に応じた制御コマンドを生成する。すなわちこの第4の実施の形態の情報処理装置4の場合のテキスト変換制御部29は、前記第2の実施の形態の場合の処理に加えて、血圧・心拍数データベース部36に予め用意されている様々な人の血圧・心拍数の特徴を現す血圧・心拍数パラメータのパターンと、上記血圧・心拍数解析部35により得られた血圧・心拍数パラメータとを照合し、その照合により話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを詳細に特定するようしている。具体的には、高い血圧や速い心拍数により「驚き」や「怒り」、「恐れ」などの感情を特定したり、低い血圧やゆっくりした心拍数により「安心」等の感情を特定する。また、血圧値や心拍数により、身長、体重、体調の善し悪し、年齢、職業、出身地などを特定して制御コマンドを生成する。なお、本実施の形態においても第2の実施の形態の場合と同様に、前記制御コマンドに応じたテキスト変換処理と前記感情等との対応付けは一例であり、当該システムで任意に設定可能であり、上記の例に限定されるものではない。

【0071】このように、本発明の第4の実施の形態によれば、入力音声信号だけでなく、例えば話者の血圧・心拍数の測定信号をも用いているため、話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などをより正確に特定可能となつてゐる。

【0072】[第5の実施の形態の情報処理装置] 図5

には、本発明の第5の実施の形態として、入力音声信号と共に例えば話者の現在位置情報等も用いて制御コマンドを生成するようにした情報処理装置5の概略構成を示す。なお、この図5の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図5の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0073】この第5の実施の形態の情報処理装置5の場合、GPS(Global Positioning System)信号入力部37には、例えば音声入力を行なっている話者の現在位置を測定するGPS測位装置からの緯度・経度信号が入力され、その緯度・経度信号はテキスト変換制御部29へ送られる。

【0074】当該テキスト変換制御部29では、前述同様に入力音声信号に基づいた話者の感情等を特定すると共に、上記緯度・経度信号を用いて特に話者の現在位置を特定し、それらの特定データに応じた制御コマンドを生成する。すなわちこの第5の実施の形態の情報処理装置5の場合のテキスト変換制御部29は、前記第2の実施の形態の場合の処理に加えて、位置データベース38に予め用意されている各地の緯度・経度情報と、上記GPS信号入力部37からの緯度・経度信号とを照合し、その照合により話者の現在位置を特定するようしている。

【0075】このように、本発明の第5の実施の形態によれば、入力音声信号だけでなく、話者の現在位置を特定するようしているため、例えば、話者が位置している場所に応じた方言や言語変換が行われる際に有用な制御コマンドを生成することが可能となっている。

【0076】[第6の実施の形態の情報処理装置] 図6には、本発明の第6の実施の形態として、入力音声信号と共に例えば話者により入力される様々なユーザ設定情報を用いて制御コマンドを生成するようにした情報処理装置6の概略構成を示す。なお、この図6の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図6の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0077】この第6の実施の形態の情報処理装置6の場合、ユーザ設定信号入力部39には、ユーザ(話者等)が例えばキーボードやマウス、携帯情報端末を操作することにより入力したユーザ設定信号が供給され、そのユーザ設定信号はテキスト変換制御部29へ送られる。

【0078】当該テキスト変換制御部29では、前述同様に入力音声信号に基づいた話者の感情等を特定すると共に、上記ユーザ設定信号を用いてさらに詳細な特定を行い、それらの特定結果に応じた制御コマンドを生成する。すなわちこの第6の実施の形態の情報処理装置6の

場合、前記第2の実施の形態の場合の処理に加えて、ユーザ設定信号入力部39からは、ユーザにより直接、話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などの情報が入力され、テキスト変換制御部29では、それらそれらの情報に応じた制御コマンドを生成する。

【0079】このように、本発明の第6の実施の形態によれば、入力音声信号だけでなく、ユーザにより直接、上記特定を行うための情報の入力が可能となっているため、前述した第2～第5の実施の形態のように入力音声信号や画像、血圧・心拍数、緯度・経度などを機器が検出する場合よりも、正確且つ確実に前記感情等の特定ができるうことになる。また、この第6の実施の形態の場合、上記感情等を特定する際に用いる情報が直接入力可能となされているため、逆に、ユーザの感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などとは全く異なる情報を自由に入力することも可能となる。言い換えると、この第6の実施の形態によれば、ユーザが自由に情報を入力可能となっているため、例えば後に前記第1の実施の形態のように、テキストデータに基づいて音声合成や言語変換等が行われることになったとき、前記第1の実施の形態の情報処理装置1に対して、当該ユーザが意図した音声合成処理や言語変換処理を行わせることが可能となる。

【0080】[第7の実施の形態の情報処理装置] 上述した第2～第6の実施の形態の情報処理装置では、入力音声信号をテキストデータへ変換し、そのテキストデータを上記制御コマンドによりさらに変換する例を挙げているが、本発明はそれらに限定されず、例えば、図7に示す第7の実施の形態の情報処理装置7のように、入力されたテキストデータに対して、上記制御コマンドによる変換処理を施すようなことも可能である。なお、この図7の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図7の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0081】この図7に示す第7の実施の形態において、テキストデータ入力部41には、例えばキーボードや携帯情報端末から入力されたテキストデータ、或いは、通信回線を介したテキストデータ、記録媒体から再生されたテキストデータなどが供給される。当該テキストデータは、テキスト変換部42へ送られる。

【0082】また、端子50には、例えば前記第2～第6の実施の形態と同様にして生成された特定結果の情報が入力されてテキスト変換制御部29へ送られる。

【0083】本実施の形態のテキスト変換部42では、上記テキストデータ入力部41から供給されたテキストデータに対して、上記テキスト変換制御部29により前述の各実施の形態と同様にして生成された制御コマンドによる変換処理を施す。

【0084】この第7の実施の形態によれば、前述の第2～第6の各実施の形態のように入力音声信号から変換したテキストデータではなく、キーボードや携帯情報端末から入力されたテキストデータ、通信回線を介したテキストデータ、記録媒体から再生されたテキストデータなど、任意のテキストデータに対して上記制御コマンドに応じた変換処理を施すことが可能となる。

【0085】[第8の実施の形態の情報処理装置] また本発明によれば、例えば図8に示す第8の実施の形態の情報処理装置8のように、手話画像をテキストデータへ変換し、そのテキストデータに対して上記制御コマンドに応じた変換処理を施すことも可能である。なお、この図8の構成は、ハードウェア構成、或いは、ソフトウェアの何れでも実現可能である。また、この図8の構成において、図2と同じ構成要素にはそれぞれ図2の場合と同様の指示符号を付し、それらの説明は省略する。

【0086】この第8の実施の形態の情報処理装置8の場合、手話画像信号入力部51には、例えば手話をしている人を撮像した動画像信号が入力され、その動画像信号は手話画像解析部52へ送られる。

【0087】手話画像解析部52では、例えば、手話をしている人のアウトラインを抽出し、次に、その人の身体の特徴点を抽出し、さらに、手話の手形状、開始位置及び動作軌跡を検出することで、手話をしている人の動作データを取得する。すなわち、当該手話画像解析部52では、手話動画像から例えば1/30秒のフレーム毎の時間差分画像を求め、その時間差分画像から手話動画像中で動きの激しい両手及び指の画像部分を抽出すると共に、それら両手及び指からなる手形状と、それら両手及び指の位置の動作軌跡を検出し、それらを動作データとして取得する。この動作データは、手話認識部53へ送られる。

【0088】上記手話認識部53は、上記手話画像解析部52から供給された動作データと、手話動作データベース部54に予め各手話の単語毎に用意されている手話動作の特徴を現す動作パターンとの照合を行い、その照合により得られた動作パターンから手話単語を求め、得られた手話単語をテキスト変換部26へ送る。

【0089】上記手話認識部53から手話単語を受け取ったテキスト変換部26は、テキストデータベース部25に予め用意されている単語モデルと、上記手話単語との照合を行ってテキストデータを生成する。

【0090】また、テキスト変換制御部29は、上記手話認識部53での認識処理による手話単語や、上記テキスト変換部26にて上記手話単語から変換したテキストデータなどに基づいて、前述同様に、手話者の感情、思考、体調、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などを特定し、それらの特定結果に応じた制御コマンドを生成する。

【0091】この第8の実施の形態によれば、前述の第

2～第6実施の形態のように入力音声信号から変換したテキストデータではなく、手話画像から求めたテキストデータに対して上記制御コマンドによる変換処理を施すことが可能となる。

【0092】[第9の実施の形態の情報処理装置] 前述の第1の実施の形態で説明した情報処理装置1は、テキストデータから合成音声を生成すると共に、上記テキストデータから抽出した所定の情報に基づいて上記加工を行う例を挙げているが、本発明は、図9に示す第9の実施の形態の情報処理装置9のよう、テキストデータから例えば手話画像を生成すると共に、上記所定の情報に応じてその手話画像の加工を行う場合にも適用可能である。なお、この図9の構成は、ハードウェア構成、或いはソフトウェアの何れでも実現可能である。また、この図9の構成において、図1と同じ構成要素にはそれぞれ図1の場合と同様の指示符号を付し、それらの説明は省略する。

【0093】この図9において、テキスト分析部11により得られた表音文字列のデータは、手話画像合成部61へ送られる。

【0094】手話画像合成部61は、手話画像データベース部62に予め用意されている手話画像辞書を用いて、上記表音文字列に対応する手話画像を読み出し、手話画像を組み上げる。

【0095】また、この図9の情報処理装置9において、加工制御部64は、情報抽出部16から供給された所定の情報に基づいて、上記手話画像合成処理やテキスト分析処理に対して加工を施す際の加工制御データを生成し、その加工制御データを手話画像合成部61やテキスト分析部11へ送る。

【0096】本実施の形態の場合、手話画像合成部61では、前述した合成音声に対する加工制御と略々同様な加工処理を、上記手話画像に対して行う。すなわち、前述した合成音声に対する加工処理のように、語尾を加工する制御や、怒り、楽しさを現す単語や語句を付加するような加工制御、言語変換(翻訳)のような加工制御等を、手話画像により実現する。なお、これら加工制御データと手話画像との対応付けは一例であって、システムで任意に設定可能であり、上記の例に限定されるものではない。

【0097】本実施の形態の情報処理装置9では、上記手話画像合成部61により合成された手話画像が、手話画像信号出力部63から図示しない後段のモニタ装置等へ送られて表示される。なお、当該手話画像を例えばネットワーク上に伝送する場合、上記手話画像信号出力部63は、上記ネットワークへの接続手段を含むことになる。また、上記手話画像信号を記録媒体に記録する場合、上記手話画像信号出力部63は、記録媒体への記録手段を含むことになる。

【0098】このように、第9の実施の形態の情報処理

装置9によれば、テキストデータから手話画像を生成すると共に、テキストデータから抽出した所定の情報に応じてその手話画像の加工を行うことが可能となっており、例えば聴覚障害を持った人でもその内容を認識することが可能となる。

【0099】[本実施の形態の情報変換処理や情報加工処理の概略ブロック構成] 図10には、上述した各実施の形態の情報変換処理や情報加工処理を実現する情報処理プログラムを実行する情報処理装置の概略的なブロック構成を示す。なお、この図10には、本実施の形態の情報処理プログラムを例えればパーソナルコンピュータにより実行することにした場合の主要部の構成のみを示している。

【0100】図10において、記憶部108は、例えればハードディスク及びそのドライブからなる。上記ハードディスク内には、オペレーティングシステムプログラムの他、本実施の形態の情報変換処理や情報加工処理をソフトウェア的に実現する情報処理プログラムや、CD-ROMやDVD-ROM等の各種の記録媒体からプログラムやデータを取り込んだり、通信回線を介して情報の送受を行いうためのプログラムなどを含む各種のプログラム109と、上記各実施の形態で説明した各データベース部として予め用意されているデータベース111、その他の各種データ110が記憶されている。なお、本実施の形態の情報処理プログラムは、上記記録媒体からインストールしたり、通信回線を介してダウンロード可能なプログラムである。また、上記データベースも記録媒体130や通信回線を介して取得可能なものであり、本実施の形態の情報処理プログラムに付随して、若しくは単独で提供されるものである。

【0101】通信部101は、例えれば、アナログ公衆電話回線に接続するためのモデム、ケーブルテレビジョン網に接続するためのケーブルモデム、ISDN(総合ディジタル通信網)に接続するためのターミナルアダプタ(Terminal Adaptor: TA)、ADSL(Asymmetric Digital Subscriber Line)に接続するためのモデムなどのように、外部とデータ通信を行うための通信デバイスである。通信1/F部102は、上記通信部101と内部バス(BUS)との間でデータのやりとりを可能とするためのプロトコル変換等を行うインターフェイスデバイスである。本実施の形態では、上記通信部101及び通信1/F部102により、例えればいわゆるインターネットへの接続、検索、閲覧、電子メールの送受信等が可能となされている。

【0102】外部機器106は、例えばテーブレコーダ、ディジタルカメラ、ディジタルビデオカメラなど、音声信号や画像信号を扱う機器である。外部機器1/F部107は、上記外部機器106から供給された信号を内部に取り込むためのインターフェイスデバイスである。

【0103】入力部113は、例えばキーボードやマウス、タッチパッドなどの入力装置であり、ユーザI/F部112は、上記入力部113からの信号を内部に供給するためのインターフェイスデバイスである。

【0104】ドライブ部115は、例えばCD-ROMやDVD-ROM、フロッピディスク等のディスク媒体130や、カード状等の半導体メモリなどから、各種のプログラムやデータを読み出し可能なドライブ装置である。ドライブI/F部114は、上記ドライブ部115からの信号を内部に供給するためのインターフェイスデバイスである。

【0105】表示部117は、例えばCRT(陰極線管)や液晶等の表示デバイスであり、表示ドライブ部116は上記表示部117を表示駆動させるドライブデバイスである。

【0106】CPU103は、上記記憶部108に記憶されているオペレーティングシステムプログラムや上記プログラム109に基づいて、図10のパーソナルコンピュータの全動作を制御する。

【0107】ROM104は、例えばフラッシュメモリ等の書き換え可能な不揮発性メモリからなり、図10のパーソナルコンピュータのBIOS(Basic Input/Output System)や各種の初期設定値を記憶している。RAM105は、記憶部108のハードディスクから読み出されたアプリケーションプログラムなどがロードされ、また、CPU103のワークRAMとして用いられる。

【0108】この図10に示す構成において、CPU103は、上記記憶部108のハードディスクから読み出されてRAM105にロードされたアプリケーションプログラムの一つである、前述した本発明の情報処理プログラムを実行することにより、前述の各実施の形態で説明した情報変換処理や情報加工処理を行う。

【0109】[情報伝送システムの構成]図11には、上述した本発明実施の形態の情報変換処理や情報加工処理の一方若しくは両方の機能を備えた情報処理装置150～153や携帯型情報処理装置(携帯電話等を含む)154、情報の配信や管理を行うサーバ161が、例えばインターネットなどの通信網160に接続された、本発明の情報伝送システムの概略構成を示す。

【0110】この図11において、本実施の形態の情報処理装置150～154等によりネットワーク上に送信されたテキストデータは、通信網160を介して直接、或いは、サーバ161の管理の元で、本実施の形態の他の情報処理装置150～154に送信される。

【0111】上記テキストデータを受信した各装置は、それぞれ、前述したように、テキストデータから抽出した所定の情報を応じて合成音声や手話画像の加工等を行う。また、本実施の形態の情報伝送システムでは、サーバ161が本発明実施の形態の情報処理プログラムやデータベース等を含む各種のソフトウェアをデータベース

162上に用意しており、各装置からの要求に応じてそれらソフトウェアを提供可能となされている。

【0112】[本発明実施の形態のまとめ]以上のように、本発明の各実施の形態によれば、情報の変換を行う場合において、例えば感情や思考、性別、顔形、身長、体重、年齢、職業、出身地、趣味、嗜好などの表現を伴った豊かで楽しい情報交換や加工を実現可能としており、また、例えば情報伝達を行う場合において、伝送情報量を増やすことなく、円滑で新しい形のコミュニケーションを実現可能となっている。さらに、本発明実施の形態によれば、例えば聴覚障害を持った人や視覚障害を持った人に対しても、新しく楽しいコミュニケーションを実現可能となっている。

【0113】なお、上述した各実施の形態の説明は、本発明の一例である。このため、本発明は上述した各実施の形態に限定されることなく、本発明に係る技術的思想を逸脱しない範囲であれば、設計等に応じて種々の変更が可能であることはもちろんである。

【0114】また、本発明実施の形態では、上述したような利点の他、情報伝達を行う場合において伝送情報量を減らすことができるため、例えば通信回線が低速回線であってもリアルタイムでの情報伝達が可能となり、また、テキストデータを記録する場合において情報量が少ないため会話や手話の内容を長時間分蓄積できる。また、それら会話や手話の内容をログ(議事録)としてテキスト形式で保存でき、さらにテキスト検索により所望の内容を検索することも可能となる。

#### 【0115】

【発明の効果】本発明は、ネットワーク等を介して受信した文字データを変換した他の情報に対して所定の加工処理を施すようにしており、その所定の加工処理として、例えば感情を表現する語を付加、若しくは感情を表現する語への変換などを行うよう正在しているため、例えば感情を伴った豊かで楽しい表現を可能とする情報交換を実現可能である。

【0116】また、本発明は、入力情報の特徴を現す情報を抽出し、入力情報から変換した文字データに対して当該抽出した特徴を現す情報に基づく変換処理を施すようにしており、入力情報の特徴を現す情報として例えば感情を表現する情報を抽出するようにしているため、文字データに対して感情を表現する情報を入れ込むことが可能となっている。

【0117】したがって、本発明によれば、例えば情報伝達を行う場合において、伝送情報量を増やすことなく、円滑なコミュニケーションを実現可能となっている。

#### 【図面の簡単な説明】

【図1】本発明の第1の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図2】本発明の第2の実施の形態の情報処理装置の概

略構成を示すブロック図である。

【図3】本発明の第3の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図4】本発明の第4の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図5】本発明の第5の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図6】本発明の第6の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図7】本発明の第7の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図8】本発明の第8の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図9】本発明の第9の実施の形態の情報処理装置の概略構成を示すブロック図である。

【図10】本発明の情報処理装置としての機能を実現するパーソナルコンピュータの構成を示すブロック回路図である。

【図11】本発明の情報伝送システムの概略構成を示す\*

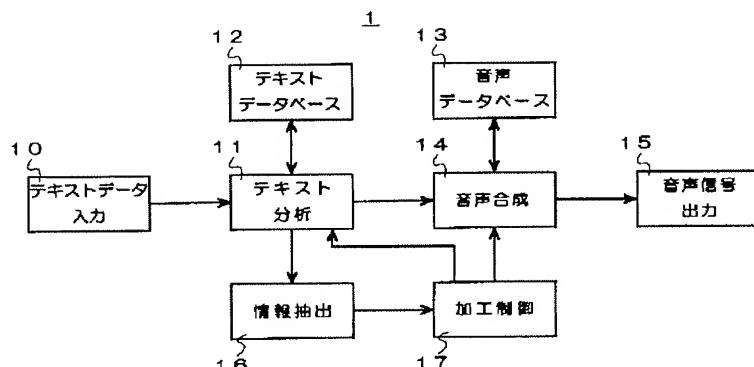
\* 図である。

#### 【符号の説明】

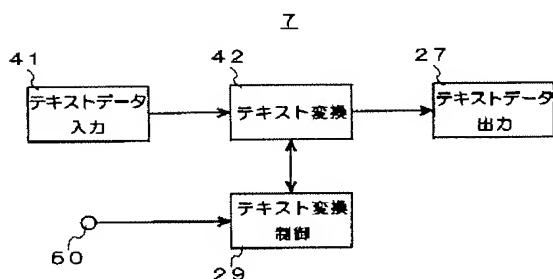
1, 2, 3, 4, 5, 6, 7, 8, 9, 150～154  
…情報処理装置、10…テキストデータ入力部、11…  
テキスト分析部、12, 25…テキストデータベース  
部、13, 24…音声データベース部、14…音声合成  
部、15…音声信号出力部、16…情報抽出部、17,  
64…加工制御部、21…音声信号入力部、22…音声  
分析部、23…音声認識部、26, 42…テキスト変換  
部、27…テキストデータ出力部、29…テキスト変換  
制御部、30…声紋・特徴データベース部、31…画像  
信号入力部、32…画像解析部、33…画像データベ  
ース部、34…血圧・心拍数信号入力部、35…血圧・心  
拍数解析部、36…血圧・心拍数データベース部、37  
…G P S 信号入力部、38…位置データベース部、39  
…ユーザ設定信号入力部、51…手話画像信号入力部、  
52手話画像解析部、53…手話認識部、54…手話動  
作データベース部、61…手話画像合成部、62…手話  
画像データベース、63…手話画像信号出力部

10

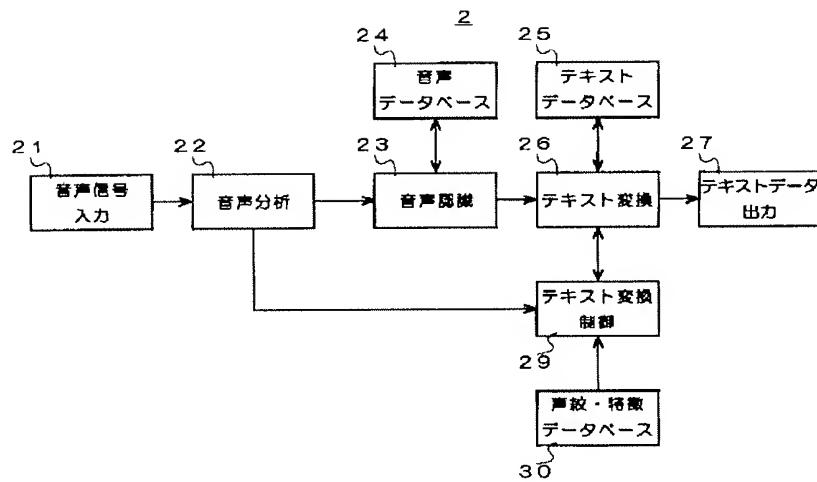
【図1】



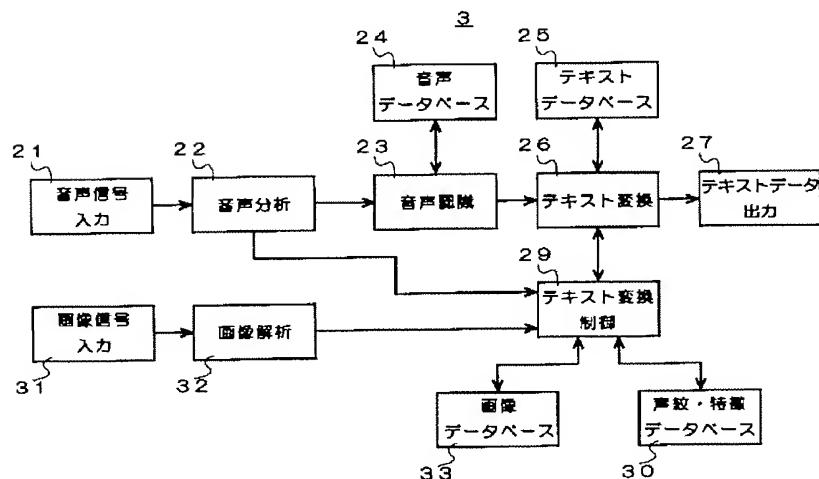
【図7】



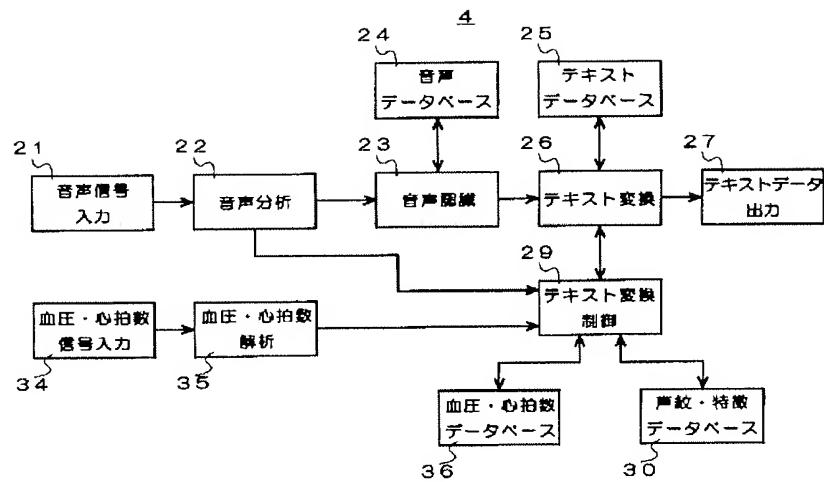
【図2】



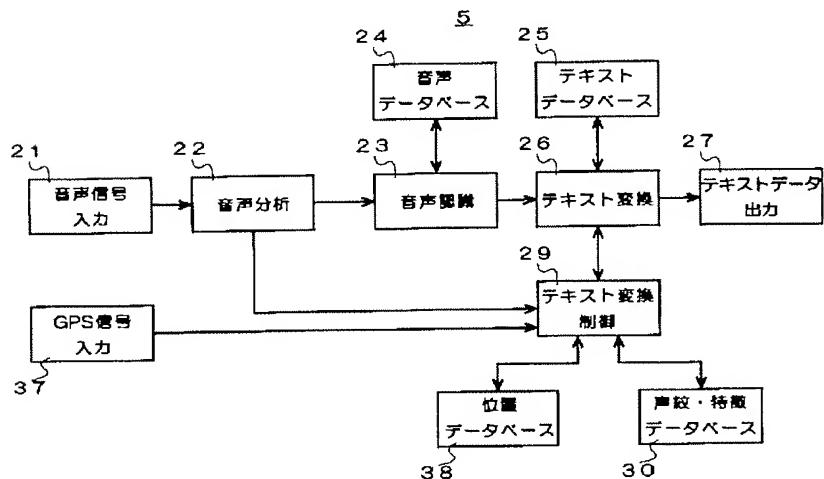
【図3】



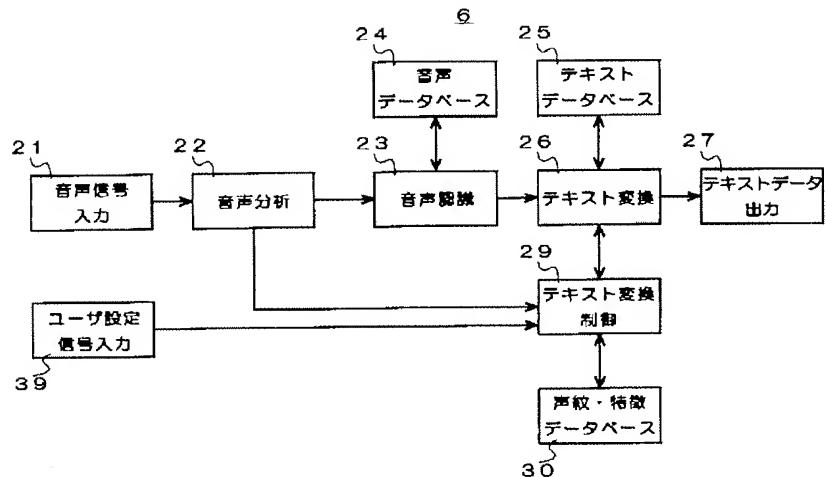
【図4】



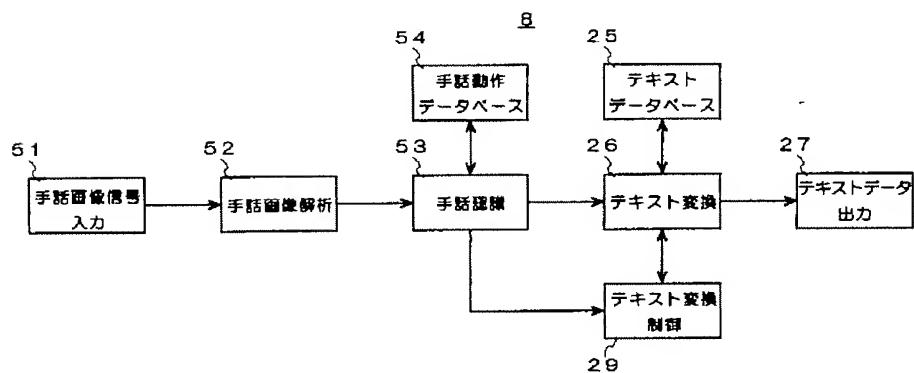
【図5】



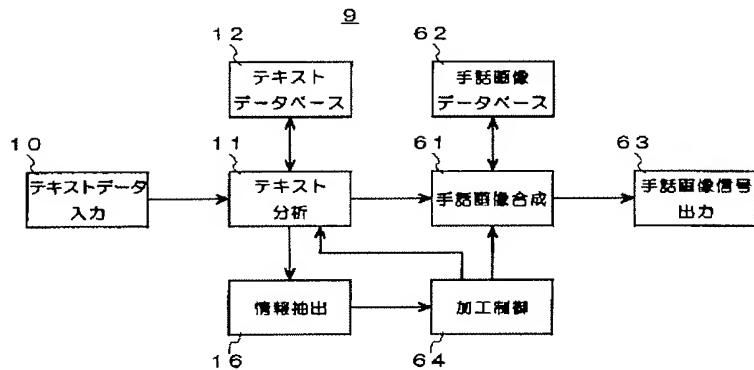
【図6】



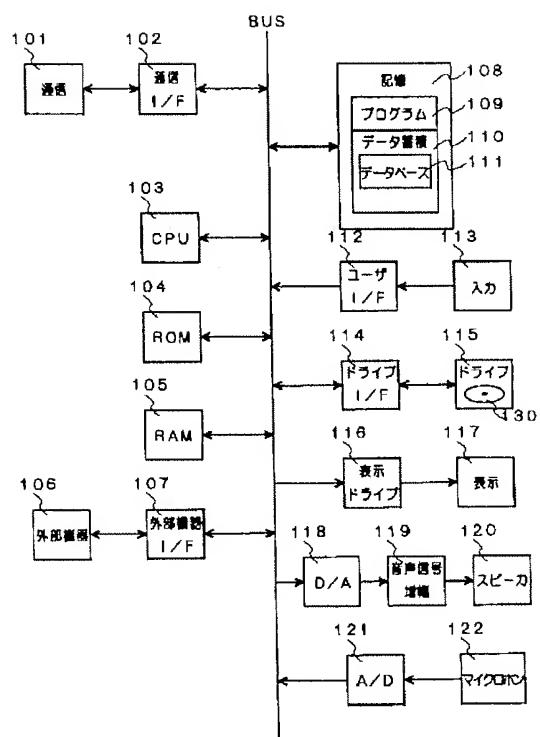
【図8】



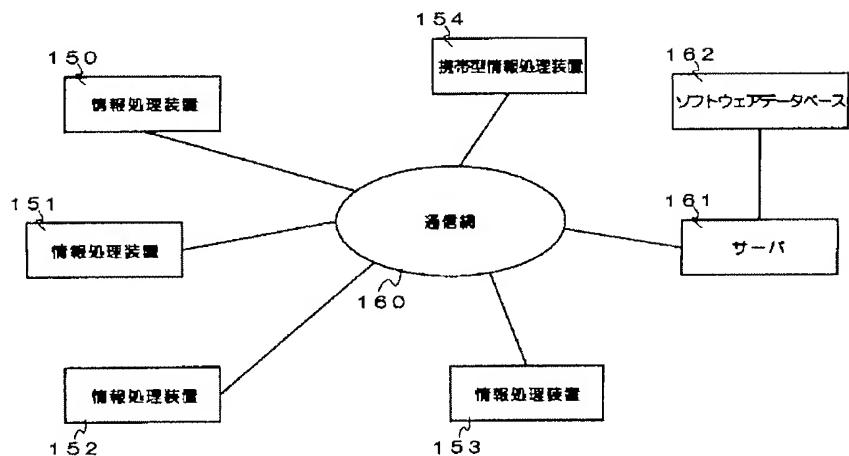
【図9】



【図10】



【図11】



フロントページの続き

(72)発明者 大平 俊充  
東京都港区赤坂7丁目1番1号 株式会社  
ソニー・コンピュータエンタテインメント  
内

Fターム(参考) 5B009 KA08 KB00 NA07 NB11 RD03  
5D015 KK02  
5D045 AA07